Preservation Planning, Trust, and Risks

Christoph Becker
http://www.ifs.tuwien.ac.at/~becker

Institute of Software Technology and Interactive Systems
Vienna University of Technology, Austria

http://www.ifs.tuwien.ac.at/dp

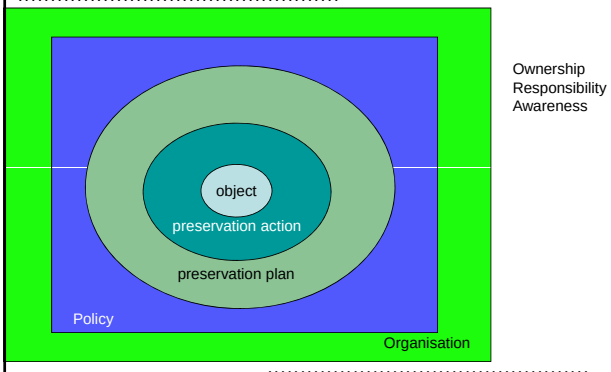
Agenda

- Preservation Planning 2: Recap, Demo, Exercise
  - What is a preservation plan
  - How to create a preservation plan
  - The planning tool Plato
- Why preservation planning?
  - Trustworthy repositories
  - TRAC
- Risk assessment: DRAMBORA
  - Risk assessment for repositories
  - Exercise: Risks in preservation planning

Definition of a Preservation Plan

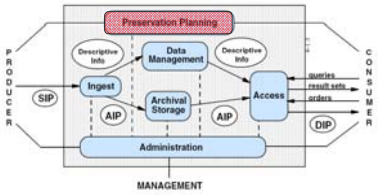
- 'A ***preservation plan*** defines a series of preservation actions to be taken by a responsible institution to address an identified risk for a given set of digital objects or records (called collection).'
- The Preservation Plan takes into account the preservation **policies**, **legal obligations**, **organisational** and **technical constraints**, **user requirements** and **preservation goals**.
- It also describes the preservation **context**, the **evaluated alternative preservation strategies** and the **resulting decision** for one strategy, including the rationale of the decision

Objects in Context

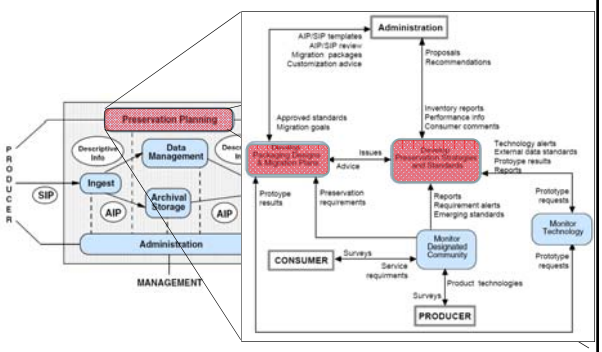


Graphic by Hans Hofman, NANETH

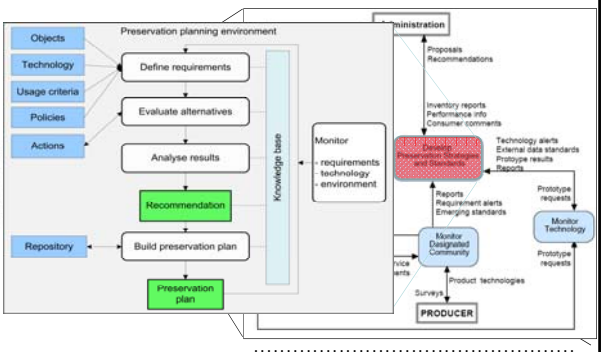
Preservation Planning and OAIS

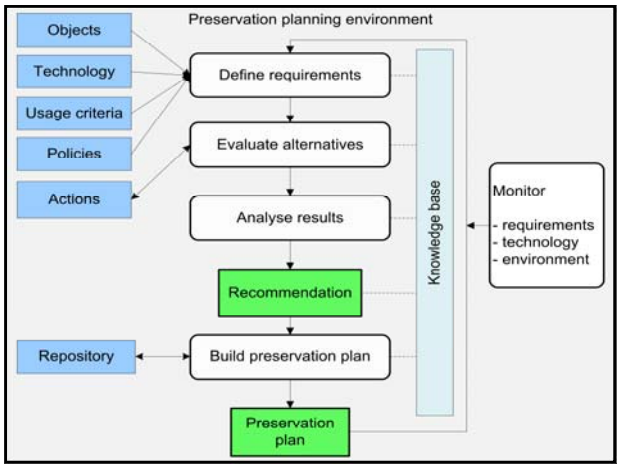


Preservation Planning and OAIS



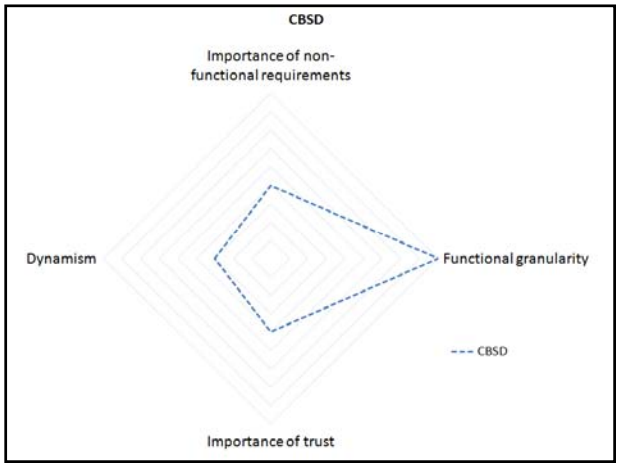
Preservation Planning and OAIS

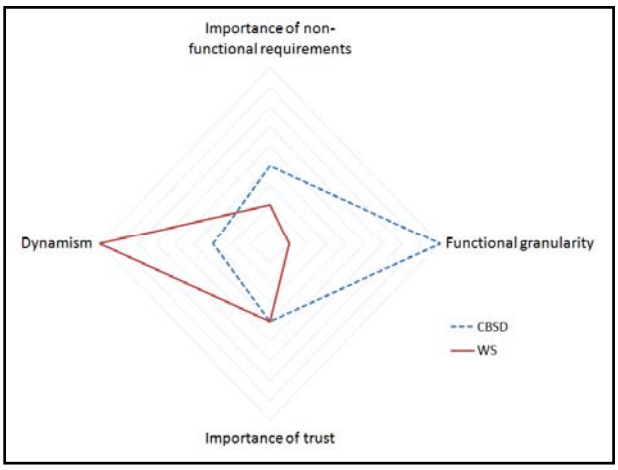


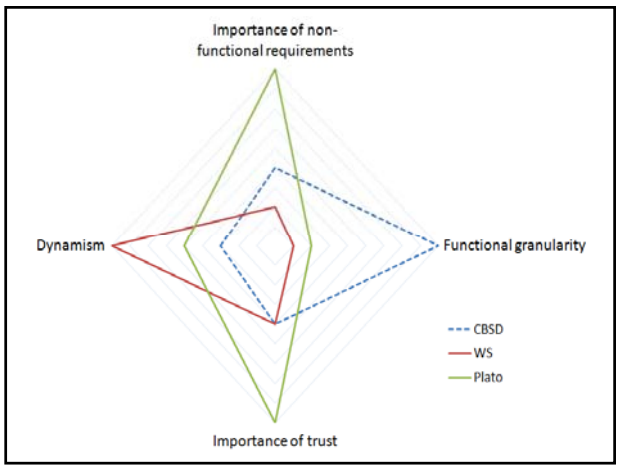


Challenges

- Several actions available (migration, emulation)
- Challenges:
  - Quality varies across tools
  - Properties vary across content
  - Usage varies across communities
  - Requirements vary across scenarios
  - Risk tolerance varies across collections
  - Preferences and constraints vary across organisations
  - Cost structures and compatibility varies across environments
  - Constraints, priorities and requirements shift
- Component selection vs. Service selection





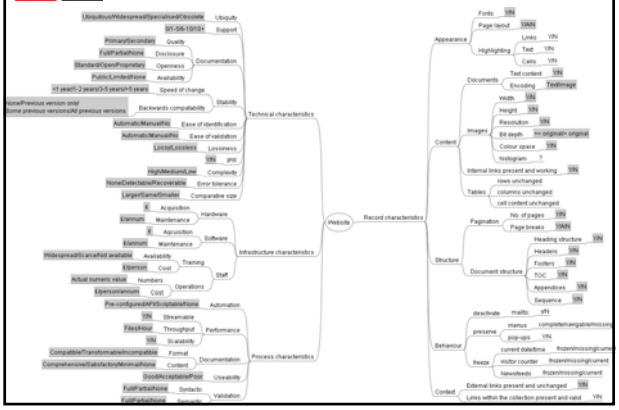


Challenges

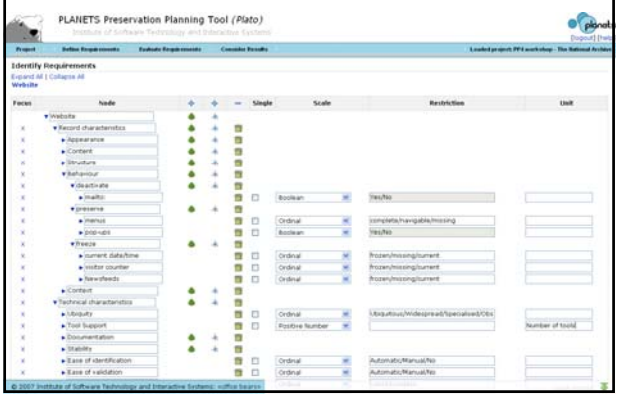
- Decision making with multiple competing objectives
- Documentation, trust, evidence, automation, scalability

Scenario	Chosen action	Main reasons
80TB scanned newspapers in TIFF5	Migrate to JP2	Storage costs, Standardisation
72TB scanned book pages in TIFF6	Leave unchanged and monitor	Color profile complications, Lack of JP2 browser support
Aerial photographs in TIFF6	Leave unchanged and monitor	Lack of JP2 browser support, Process costs

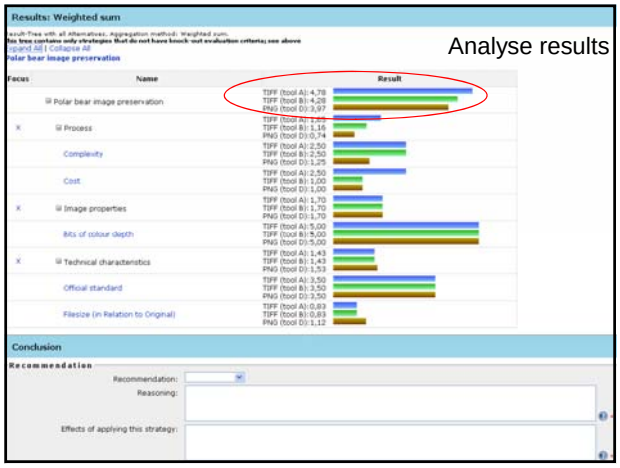
An Objective Tree

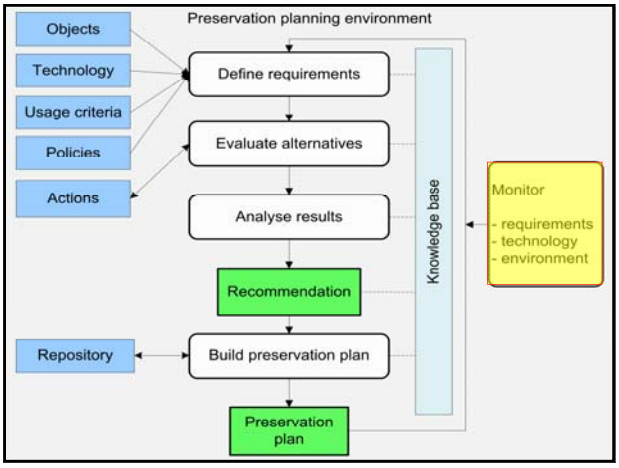


Objective Tree in Plato



Analyse results





Conclusions

- Methodologically sound model to specify and document requirements
- Repeatable and documented evaluation for informed and accountable decisions
- Set of templates to assist institutions
- Generic workflow that can easily be integrated in different institutional settings
- Plato
  - Tool support to perform solid, well-documented analysis
- Provides basic preservation plan

http://www.ifs.tuwien.ac.at/dp/plato

Questions?

Demo Time...

www.ifs.tuwien.ac.at/dp/plato

- Exercise time!
- www.ifs.tuwien.ac.at/dp/plato

- dpvo1-4
- (Create and) load „DEMO PLAN scanned images“
- Analyse provided documentation
- Discuss evaluation+transformation
- Analyse and draw conclusions

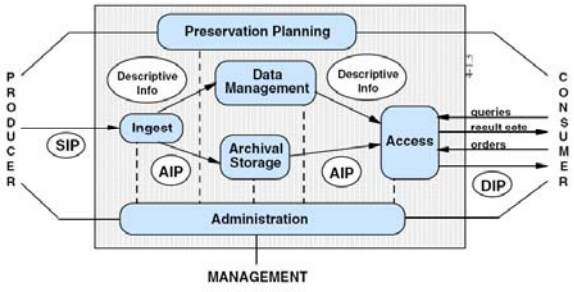
Why all that? .... Trust and Risks

- Trust in Digital Repositories
- TRAC
- Nestor Criteria
- Risk and Digital Preservation
- Principles of DRAMBORA
  - Overview
  - Workflow
  - Results
  - Benefits
- Exercise: Risk assessment of preservation planning

Trustworthy repositories

- Producers and consumers need trust in a repository
- What is trust?
- Concepts
  - being able to predict something
  - Confidence of producers and consumers
  - Reliability, authenticity
  - A trusted party is presumed to seek to fulfill expectations (legal obligations, policies, ethics, contracts…)
- Standards
- OAIS compliance...?

The OAIS Model



Trust in a Repository

- Critical services require trust
- RLG/OCLC “Trusted Digital Repositories – Attributes and Responsibilities” (2002)
  - depositors trust information holders
  - users trust digital assets provided by repositories
  - information holders trust third party service providers
- How is trust established, maintained, and secured?
- How to verify trust?

Trustworthy Repositories Principles

CRL-RLG-OCLC-Nestor-DPE-DCC criteria and checklists

I. The repository commits to continuing maintenance of digital objects for identified community/communities.

II. Demonstrates organizational fitness (including financial, staffing structure, and processes) to fulfil its commitment.

III. Acquires and maintains requisite contractual and legal rights and fulfils responsibilities.

IV. Has an effective and efficient policy framework.

V. Acquires and ingests digital objects based upon stated criteria that correspond to its commitments and capabilities.

Trustworthy Repositories Principles

VI. Maintains/ensures the integrity, authenticity and usability of digital objects it holds over time.

VII. Creates and maintains requisite metadata about actions taken on digital objects during preservation as well as about the relevant production, access support, and usage process contexts before preservation.

VIII. Fulfils requisite dissemination requirements.

IX. Has a strategic program for preservation planning and action.

X. Has technical infrastructure adequate to continuing maintenance and security of its digital objects.

Audit and Certification Initiatives

- RLG- National Archives and Records Administration Digital Repository Certification Task Force
  - Trustworthy Repositories Audit & Certification: Criteria and Checklist (TRAC)
- NESTOR
  - Catalogue of Criteria of Trusted Digital Repositories
- DRAMBORA: Self-assessment

TRAC

Criteria checklist

Three groups

A. Organisational Infrastructure

B. Digital Object Management

C. Technologies, Technical Infrastructure & Security

TRAC and Preservation Planning I

A 3.2 Repository has procedures and policies in place, and mechanisms for their review, update, and development as the repository grows and as technology and community practice evolve.
- Planning procedure
- Watch Services, triggers
- Update of preservation plans

A3.6 Repository has a documented history of the changes to its operations, procedures, software, and hardware that, where appropriate, is linked to relevant preservation strategies and describes potential effects on preserving digital content.
- History of preservation plans (created, reviewed and updated)
- Plato: Automated documentation of planning activities

TRAC and Preservation Planning II

A3.7 Repository commits to transparency and accountability in all actions supporting the operation and management of the repository, especially those that affect the preservation of digital content over time.

- Solid workflow in consist manner enables informed and well-documented decisions
- Explicit definition of objectives and measurement units
- Change history in plans

B1.1 Repository identifies properties it will preserve for digital objects.

- Objective Tree
- Evaluation results

TRAC and Preservation Planning III

B3.1 Repository has documented preservation strategies.

- Preservation Plan

B3.3 Repository has mechanisms to change its preservation plans as a result of its monitoring activities.

- Watch Services, triggers
- Verification against changes in the environment
- Update of preservation plans

Nestor Criteria & Preservation Planning

8. The digital repository has a strategic plan for its technical preservation measures.

- Preservation Plan
- Triggers for re-evaluation
- Watch Services

9.2 The digital repository identifies which characteristics of the digital objects are significant for information preservation.

- Objective Tree

- Cf. TRAC B1.1!

Nestor and TRAC versus PP

- Certification and Audit of repositories
- NESTOR and TRAC
- TRAC ISO certification in progress

- Planets Preservation Planning approach
  - Documented preservation strategies
  - Identification of significant properties
  - Continuous monitoring and mechanisms to react to changes in the environment

...and in practice?

- Criteria checklists important step
  - Future: audit certificates

- Criteria not always helpful
  - How to measure fulfilment
  - How to prove trust
  - How to improve

- Audit and Certification as ultimate goal
- Self-audit as important step

DRAMBORA

- Digital Repository Audit Method Based on Risk Assessment
- Self-Audit and Self-Assessment

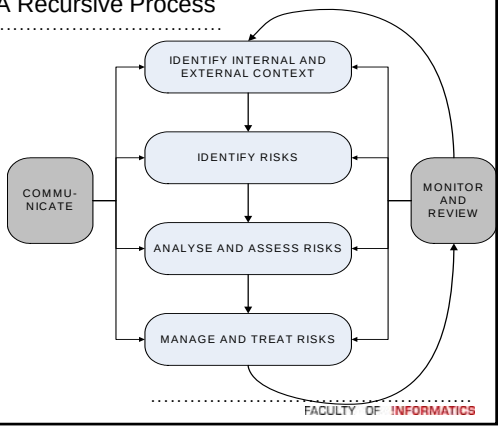
- Evidence based
  - Consistency
  - To ensure conclusions can be validated and replicated
  - Documentary, testimonial, and observational evidence

- Pilot audits
- Risk awareness is low within the community

Risk and Digital Preservation

- Digital Preservation ***is*** Risk Management
- Transform uncertainties into manageable risks
- Standard risk management models in many disciplines
- DRAMBORA is adaption of standard risk assessment procedure, customized to DP

A Recursive Process



DRAMBORA workflow

1. Define the audit scope
2. Document the context
   - Determine functional classes to organise activities and assets
3. Formalise and document the organisation
   - Mandate, constraints
   - Goals, objectives
   - Activities
   - Assets
4. Identify, assess and manage risks

Organisational Goals

- Associated with one of 8 functional classes
  - Acquisition & Ingest
  - Preservation & Storage
  - Metadata Management
  - Access & Dissemination

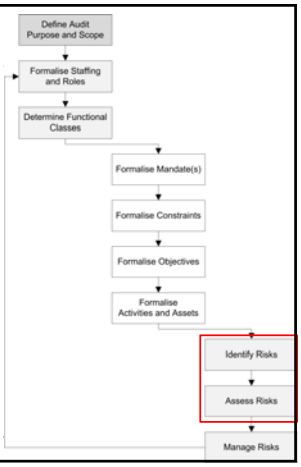
  } operation classes

  - Organisation & Management
  - Staffing
  - Financial Management
  - Technical Infrastructure & Security

  } supporting classes

Risk assessment

- Start with activities and assets
- Identify vulnerabilities and threats
  - Assets or activities may fail to achieve relevant goals
  - Internal and external threats may pose obstacles to the success of activities
- Assess risks
  - Probability
  - Impact
  - Relationships between risks



Organisational Assets

- Includes:
  - information (databases, data files, contracts, agreements, documentation, policies and procedures)
  - software assets
  - physical assets
  - services and utilities
  - processes
  - people
  - intangibles, such as reputation

Evidence

- Documentary
  - Mission statement
  - Deposit agreements
  - Business plan, financial reports
  - Job descriptions/profiles
  - System manuals, Technical documents,…
- Testimonial evidence
  - Highlight whether omissions exist in documentation
  - Validate documentation vs. reality
- Observation of practice
  - Less objective, but important
  - Walkthroughs, test objects,…

Risk assessment

- Assess risks
  - Probability
  - Impact
  - Relationships between risks
- Manage risks
  - Mitigation
  - Avoidance
  - Acceptance
- Iterative process

Risk relationships

Risk Relationship	*Definition of Risk Relationship*
Explosive	where the simultaneous execution of *n* risks has an impact in excess of the sum of each risk occurring in isolation
Contagious	where a single risk's execution will increase the likelihood of another's
Complementary	where avoidance or treatment mechanisms associated with one risk also benefit the management of another
Domino	where avoidance or treatment associated with a single risk renders the avoidance or treatment of another less effective
Atomic	where risks exist in isolation, with no relationships with other risks

Sample risk 1

Risk Identifier	**R05**
Risk Name	Repository loses mandate
Risk Description	Basis for repository's existence is withdrawn or substantially altered, rendering it incompatible with business activities
Is this Risk relevant?	Is the mandate subject to ongoing review? Is the primary repository service contract subject to renewal or renegotiation?
Example Manifestation	Scope of repository responsibility is changed by legislative amendment
Nature of Risk	Personnel, management and administration procedures
Probability	2
Potential Impact	4

Sample risk 1: Mitigation

Avoidance	Seek all available certifications to publicly demonstrate operational effectiveness
	Promote organisational transparency
In the event of execution	Establish arrangements for succesion
	Establish contingency plans
	Establish exit strategy

Sample risk 2

Risk Identifier	**R66**
Risk Name	Preservation strategies result in information loss
Risk Description	Exposure of an archived object to preservation plans result in loss or damage to one or more of its significant characteristics
Is this Risk relevant?	Does repository offer a definition of acceptable loss that may result from preservation activities?
Example Manifestation	Migration strategy results in loss of 'look and feel' of archived documents, regarded as essential properties by user community
Nature of Risk	Operations and service delivery
Probability	4
Potential Impact	3

Sample risk 2: Mitigation

Avoidance	Evaluate preservation strategies in controlled environment prior to execution
	Ensure procedures are reversible in the event of unexpected or inappropriate results
In the event of execution	Define policies to describe the acceptable levels of loss tolerated by the repository

DRAMBORA Outcomes

- Documented organisational self-awareness
- Catalogued risks
- Understanding of infrastructural successes and shortcomings
- Preparation for full scale external audit

Exercise

- Organisation: National library
- Mandate: Preserve the newspaper collections
- Situation: They are all scanned
- Activities: Preservation planning, Migration, QA

- Assets? Roles? Owners?
- Threats? Risks?
- Risk probabilities? Impact?
- Mitigation? Avoidance? Acceptance?

Thank you for your attention.

7.6.: Guest lecture TREVENTUS book scanner!

http://www.ifs.tuwien.ac.at/~becker